Ignorance in context: The interaction of modified numerals and QUDs

Matthijs Westera — ILLC, University of Amsterdam

Adrian Brasoveanu — UC Santa Cruz

I. The phenomena. Geurts & Nouwen (2007) report a contrast between superlative modifiers (henceforth SUPs) like *at most*, and comparative modifiers (henceforth COMPs) like *less than*: only SUPs license an **ignorance inference**. For instance, (1a) but not (1b) conveys that the speaker doesn't know the exact number of diamonds:

1. a. SUP: I found at most ten of the diamonds under the bed. ⇝ *not sure how many*
b. COMP: I found less than ten of the diamonds under the bed. $\not\rightsquigarrow$ *not sure how many*

This contrast was experimentally confirmed by Geurts et al. (2010), but disconfirmed by Coppock & Brochhagen (2013), who use a different experimental setup. This paper builds on their work in search of an explanation for this discrepancy. We report the results of 2 experiments, each consisting of an acceptability-judgment-like task paired with a self-paced reading task, investigating the dependence of such ignorance inferences on **questions under discussion (QUDs)**.

Experiment 1 considered (1a,b) as responses to the 3 kinds of questions in (2) below ($2 \times 3 = 6$ conditions total):

2. a. POLAR: Did you find {at most / less than} ten of the diamonds under the bed?
b. WHAT: What did you find under the bed?
c. HOWMANY: How many of the diamonds did you find under the bed?

We found an overall weaker ignorance inference in response to a POLAR (2a) than in response to a WHAT question (2b), but with no difference between SUPs and COMPs. The contrast in ignorance between SUPs and COMPs is detectable only in responses to HOWMANY questions (2c), with SUPs exhibiting stronger ignorance inferences than COMPs. *Stronger* ignorance inferences are also systematically correlated with *increased* reading times (RTs): the WHAT and HOWMANY & SUP conditions are associated with significantly higher RTs.

The follow-up Experiment 2 considered three additional question types:

3. a. APPROX: Approximately how many of the diamonds did you find under the bed?
b. EXACT: Exactly how many of the diamonds did you find under the bed?
c. DISJUNCT: Did you find eight, nine, ten, or eleven of the diamonds under the bed?

We found weaker ignorance inferences for APPROX (3a) relative to EXACT (3b) and DISJUNCT (3c), with no difference between SUPs and COMPs for any QUD type and no significant effects on RTs (except for the DISJUNCT condition in one region; more details in the paper). Below, we first present our account, and then discuss the experiment in more detail.

II. The account. We take our findings to argue for a *pragmatic* account, but *with a twist*, to account for the contrast between SUP and COMP in response to the HOWMANY QUD. We assume: **(i)** ignorance is derived via the Maxim of Quantity, in line with Grice (1975), Büring (2008): if the context is understood as demanding precision, then giving merely an upper bound implicates ignorance; **(ii)** HOWMANY is *underspecified* with regard to the desired level of precision, unlike POLAR/APPROX (imprecise) and WHAT/EXACT/DISJUNCT (precise); **(iii)** SUPs are used in contexts demanding precision more than COMPs are (Cummins et al 2012). Now, if the QUD *fixes* the desired level of precision (as do all QUDs except for HOWMANY), then the assumed difference between SUP and COMP cannot affect the ignorance inference. However, if the QUD *doesn't* completely fix the desired level of precision (as is the case for the HOWMANY QUD), participants use the modified numeral as a cue to disambiguate the context: if the answer contains a SUP, the context was likely precise, yielding a stronger ignorance inference; but if the answer contains a COMP, the context was likely less precise, yielding a weaker ignorance inference.

Our account suggests an important way in which a lexical entry's *typical context of use* affects pragmatics, by affecting how an underspecified context is disambiguated. QUD underspecification, we argue, is the reason for the discrepancies between previous experiments.

III. Experimental method and statistical modeling. Both experiments had a 2×3 design, with 36 items each (6 per condition), presented as conversations between a judge and a witness. The instructions mentioned the witness had nothing to hide. Each item consisted of a **question** by the judge, the witness's **answer**, and an **inference** drawn by the judge:

4. The judge asks: "What did you find under the bed?" (Example from Exp. 1)
The witness answers: I found at most ten of the diamonds under the bed.
Based on this, the judge concludes: "The witness doesn't know exactly how many of the diamonds she found under the bed." How justified is the judge in drawing that conclusion?

Question, answer and inference were presented on three distinct screens, with the answer presented as a **self-paced reading** task: participants read it word-by-word, with the SPACE bar revealing the next word and hiding the preceding one. Participants were asked to indicate on a 5-point Likert scale how justified they thought the judge's inference was (1: not justified at all, 5: strongly justified). We used a Latin square design for each experiment (6 lists of items, participants were rotated through lists, every item appeared once in each list with items balanced across conditions, the lists rotated the items through the conditions). There were 108 stimuli total (36 items + 72 fillers) in each experiment, the order of which was randomized for every participant (35 participants in Exp. 1, 51 participants in Exp. 2).

For reasons of space, we discuss only the statistical modeling of the acceptability (justifiability) data. Since the response data for both experiments was ordinal (ordered categorical), we used mixed-effects ordinal probit regression models to analyze the data. All models included intercept random effects for participants and items.

In Experiment 1, there were two factors: ModNum with two levels Comp (the reference level) and Sup, and QudType with three levels Polar (reference level), What and HowMany. The interaction model with all two-way interactions did not significantly reduce deviance compared to the main-effects-only model (LR statistic 2.72, $p = .26$), but when we examined subsets of the data by QudType, there was a significant effect of Sup for the HowMany subset ($\beta = .27, SE = .11, p = .016$), but not for the Polar or What subsets. In the main-effects-only model estimated for the entire data set, the main effects for both What ($\beta = .23, SE = .08, p = .003$) and HowMany ($\beta = .28, SE = .08, p = .0004$) were highly significant, but not the main effect for Sup ($\beta = .11, SE = .06, p = .08$).

The analysis of the Experiment 2 data proceeded in a very similar way (with Approx as the reference level for QudType). Once again, the interaction model with all two-way interactions did not significantly reduce deviance compared to the main-effects-only model (LR statistic 2.57, $p = .28$). Furthermore, Sup was not significant in any of the three QudType subsets. In the main-effects-only model estimated for the entire data set, the main effects for both Exact ($\beta = .23, SE = .06, p = .0003$) and Disjunct ($\beta = .17, SE = .06, p = .007$) were highly significant, but not the main effect for Sup ($\beta = .07, SE = .05, p = .21$).

References:

• Büring, D. 2008. The least *at least* can do. 26th WCCFL.
• Coppock, E. & T. Brochhagen. 2013. Diagnosing truth, interactive sincerity, and depictive sincerity, SALT 23.
• Cummins, C., U. Sauerland & S. Solt. 2012. Granularity and scalar implicature in numerical expressions, *L&P* 35.
• Geurts, B., N. Katsos, C. Cummins, J. Moons, L. Noordman. 2010. Scalar quantifiers: logic, acquisition, and processing, *Language and Cognitive Processes*, 25.
• Geurts, B. & R. Nouwen. 2007. *At least* et al.: the semantics of scalar modifiers, *Language* 83.